Referee #2

General Evaluation

This manuscript represents a field investigation of three different acoustic-based methods used to quantify the bedload transport rate over the single cross-section. It is admiring how carefully the experiments were performed and the amount of collected data is rather unique. The reading of the manuscript was insightful and brought some details into my attention that I was not aware of before, mostly regarding the passive acoustic techniques. I read the manuscript with great interests, therefore quite a few comments and discussion issues are presented in this review. The relation between the bedload dynamics and the measurement capabilities of each technique is well-addressed and by that, this article represents its fair novelty in this research field. The presented results are clear and presented in fair quality figures and the overall structure is satisfying. Besides, some appendices offer an overview of the available instrumentation and the average measured data.

Yet, it must be noted that the discussions, conclusions, and some of the methods used are not up-to-date and deserve some more thorough investigation. Obviously, the later does not refer to the hydrophone technique, but mostly to the adcps and the physical bedload sampler.

The language and the sentence structure is simple and understandable but the text, in general, deserves a brush-up. Honestly, I did not focus on that so please re-read and pay more attention to the wording.

I will write my review by giving general comments that address the different techniques and the discussion of the results. Later in the specific comment section, some less important issues will be listed, sometimes referring to the general comments (g.c).

Finally, I would like to apologize for eventual missing citations, some language unclarity or partial information. I wrote this review as quickly as I could. Please keep in mind that the paper could be published with only minor revisions, specifically in the part of the discussion and not all my comments have to be considered. However, I do think it will necessarily increase the quality of the paper.

General Comments

1. Study site

Although there are nice tables in the appendices I am missing a Table in the main text with the general hydraulic conditions and sediment characteristics.

a) In the text, there is no info about the **standard deviation of the particle's sizes** of the bedload material. I would also like to see the range of the **particle Reynolds number (Rep)**, the **shear velocity or stress and the critical shear velocity or stress**. It could be also interesting if the authors give information about the reference shear stress at the study site (Parker et. al 1982). Note that this is similar to the critical shear stress but not the same.
b) There are quite a few comments about the measurement capability of the instruments related to the different PSD of the bedload particles, but not details about the PSD of the sampled bedload surface.
c) I have another question about the boat/floating structure given ion Fig2b:

How was It fixed and how intensive was the drifting that you mention several times (in terms of meters)?

2. BTMA

A large uncertainty is associated with these physical samplers (Frings, et al., 2017). All physical sampling techniques suffer from difficulties and problems during the measuring procedure, such as: i.) disturbance of the bed; ii.) gap effect (sampler mouth is not in contact with the bed) iii.) sampling time; iv.) positioning above bedforms; v.)loss of material during sampling and raising the sampler, vi.) clogging of the mesh; etc. (Gaweesh, et al., 1994). Concerning these disadvantages, I have several questions:

a) I was wondering how this calibration coefficient (α=2) was adapted. Does it take in consideration all the factors numbers above(check (Banhold, et al., 2016) for detailed analysis of a very similar sampler in Germany) ? It would be good to see a sentence regarding this issue. Calib coefficient of means that the samples underestimate 100 % of the real sediment transport.
b) Is there a defined procedure for sampling? How many samples are sufficient to minimize the uncertainty (Frings, et al., 2017)?
c) Based on which tests? I could not find the citation online…Boiten 2003? It looks like lecture notes, so inside it must be stated how these coefficients are defined and related to another study.
d) And if some experiments exist, were those conducted in similar conditions as your study site?
e) What are the conditions when the sampler would malfunction (e.g. weak transport conditions or too abundant)? There are many examples of low transport rates could be underestimated or overestimated by the physical samplers. Although most of the literature suggests that in weak transport conditions the samplers would over-estimate (e.g., due to digging in the bed), sometimes in sandy rivers, we have noticed many samples to under-sample the total transport (observing videos and total mass caught).
f) Given the Eq. 1 I was wondering if you have done PSD and weight/sieving analysis in the laboratory for each sample? If so, why not using only the dry mass?
g) beside the sieving, there is no information on how you have measured the other parameters, such as the volume V.
h) In the appendices, you give information on how many samples are averaged in the given value. But there is no explanation of how the samples from the two BTMA are used in the processing. Please give information about that.

This information is crucial because it correlates directly with some of the conclusions regarding the measurement capacities of acoustic instruments.

3. ADCPs

In general, the methodology and the conclusions given for the ADCPs, as active sonars are not up-to-date and some of the conclusions are partially correct. Here are some comments:

a) The **post-processing procedure** only partially follows some of the latest findings given in (Conevski, et al., 2019) and (Conevski, et al., 2020). Although the authors tried to implement some of the procedure there are some miss-steps. The **de-spiking and the filtering procedure** (discarding the raw BT velocities that are in the opposite direction of the flow) given in these two studies improved the correlation with the samples and gave more realistic velocities. It seems that the filtering procedure automatically excluded the negative velocities the occur in the so-called recirculation zone of the dunes (which you also mention). In addition, there is some sort of final filtering the involves analysis of the difference of the depth values registered by each beam, which helped to discard some samples that suffered from beam inhomogeneity (Conevski, et al., 2020).

b) The **projection to the water direction** (Wdir GPS) is not clear to me. How these two are defined do you use compass or GPS heading for Wdir and compass heading for Va? Why is this necessary?
You say "To avoid compass and GPS issues, and to eliminate the effect of residual lateral displacement", but then you use again: wdir GPS (GPS ref coordinate system)-bdir BT (compass heading EH). By doing this some of the issues given above are inherited again in the velocity. Also, it is not true that the bedload velocity should align with the water velocity especially in presence of bedforms. If one would like to project one velocity vector on another both should have then the same reference coordinate system or to contain information (rotation and /or translation) about the different reference coordinate system. As far as I understand you use GPS-ENU and Compass – ENU. Note that you are in a fixed position and you do not need water velocity referenced to the BT. Another thing is that if you project over the water velocity the real direction of the bedload is lost, and the final value decreased.
Please correct me if I understood wrong.

c) The concept of volume and surface scattering seems not to be considered in conceptualizing the discussions. Please check (Urick, 1983), also (Conevski, et al., 2019) , characteristics (Conevski, et al., (2020, accepted)). Note that the scattering of the surface formed by the immobile particles below the active layer has large impedance and therefore the echo coming from this scattering is dominant. The roughness of this surface could induce false velocity estimation. This error depends on the acoustic parameters of the instrument and the signal processing algorithm. Please have in mind that this is a complex two-phase (surface +volume) scattering process occurring between two scattering regions (Rayleigh + geometric), so it cannot be easily described.

d) In the same manner, the concept of internal processing and signal processing is not well addressed. Note that the RDI uses broadband codded pulse technology while M9 uses the pulse-coherent method with some internal modifications such as the use of HD smart pulse. Note that the RDIs BB is able to use much finer resolution systems (Brumley, et al., 1991) in the detection of the bottom signal, thus it should be able to define the better immobile bed surface. Note that the main purpose of the bottom tracking signal is to identify the riverbed and it is developed in that way (R. Lee Gordon, 1996). This is one of the reasons why RDI continuously reports lower velocities than M9, regardless of the frequency (Conevski, et al., 2020). This is also related to the acoustic sampling effect described in the same paper.

e) Although it is well known that the grain size is correlated with the backscattering strength even with the bedload particles (Conevski, et al., (2020, accepted); Shiel, 2010), the laboratory investigation (Conevski, et al., 2019; Conevski, et al., (2020, accepted); Conevski, et al., 2020) demonstrated that it is not the only reason for different Doppler / apparent bedload velocities especially in well-developed transport conditions.

f) In the results section, there is no distinction in the Figures between the data measured by the RDI 1200kHz and the M9 (also which frequency). In Latosinski et .al. 2017 and Conevski et al.

2020 it is clearly indicated that different ADCPs give different mean apparent bedload velocities, and it is not strictly dependent on frequency, but also to the acoustic sampling and resolution, pulse length, cell-profiling, signal processing etc. The laboratory tests are given in Conevski et. al 2019 and (Conevski, et al., 2020) also confirm and elaborate on these issues.

g) The timing of 3-10 minutes is enough to get a stable apparent velocity at a single position (Conevski, et al., 2019; Rennie, et al., 2002).

h) What do you mean by site-specific? This is partially true, and it is only dependent on the shape of the bedforms and the riverbed surface. Some of these elements could be identified and eliminated by filtering procedure, beam inhomogeneity analysis, surveying several cross-sections before positioning. Conevski et. al (2020) reported a relatively high correlation coefficient with data collected at several different positions and two different rivers. As given in the comments above the ADCPs are more instrument dependent than site-specific.

i) The water bias could be assessed by checking the raw water velocity and backscattering echo values from the last available cell, together with the camera data. Also one should expect rather high suspended sediment concentration in the entire profile (above 500 mg/l) to have the effect of water bias (Rennie, et al., 2002; Gaeuman, et al., 2006). Although this effect has not been investigated in laboratory conditions it is expected that the long BT pulse contains more energy and therefore the signal penetrates this suspended layer. Once it penetrates the scattering reflected both, from the active bedload layer (which has an absolutely higher concentration than the SSC right above) and the surface scattering of the immobile particles below, is absolutely stronger and easier to detect in the autocorrelation as part of Doppler estimation methodology.
What was the SSC concentration in some of the cases? Could you upload a video of the sampling area?

j) The kinematic model.
<u>Shear velocity</u>. The semi-empirical approach analytically derived by VanRijn is fairly valid as long as the roughness is assumed to be 3D90. Since you have the ADCP right above the dune a log-law fitting might effectively correct or the Reynolds method could be also applied (Guerrero, 2011). By obtaining the velocity profile right above the dune stoss side, the drag effect from the bedforms might be negligible, thus the shear stress would be only related to the motion of the grain. Not sure if this is feasible in your case, but it is an idea.
It seems that the formulas you use for the critical shear velocity do not involve the hiding effect function. Note that you have sand and gravel in the bed mixtures, and therefore a fractional transport. This might not be crucial for only determining the active layer thickness but it is worth checking because the active layer thickness might increase due to the decrease of the critical shear velocity of the gravel particles, in a presence of sand (Wilcock, et al., 2001; Curran, et al., 2005).
<u>Porosity</u>. Which value do you use for the porosity of bedload concentration in the Eq. 4? A constant bedload concentration of 0.65 is utterly false. Frozen-core bed surface or substrate samples were taken from riverbed have a porosity of 0.4, which simply means the concentration of 0.6. in the same context perfectly, packed particles without consistent shape hardly reach porosity less than 0.3. In the kinematic model, the concentration value is the value of the concentration in the instantaneous (dynamical) bedload active layer. This value shall be at a

maximum lower than 0.6; Van Rijn (1984) reports this as maximum bedload concentration. Rennie and Villard (2004) reported 0.1-0.15 in their model. Note that this is not true for the porosity of the dunes; it is also stated in the paper that you make a distinction between the "dynamical active layer" (I am calling it instantaneous, or max saltation height) and the event - scale exchange layer (e.g., this includes the height of the dunes)

4.Dune tracking method

a) The river morphology and the presence of bars (that actively migrate) suggest the dune formation throughout this cross-section would be 3-dimensional, which means that I would not expect straight line dune wavefront. Is it correct?
b) On the other hand, the DTM method that you use is built based on well-developed 2D dunes. Could you please elaborate on this issue? I am afraid that maybe only the dune measurements in the thalweg zone would be most reasonable to use in the comparison with the other two techniques?
c) Why not compare the bedload transport integrated over the entire cross-section from the hydrophones and ADCPs vs the dune tracking of only several longitudinal profiles (e.g., close to the thalweg).

5.Hydrophones

a) What is the advantage of drifting of the hydrophones? Isn't there less ambient noise (not bedload flux noise) if they stay fixed? Ok, I read in the Geay et al. (2020) that it is actually the opposite… I did not find a comparison of both cases. How was drifting performed and why the waves hitting the boat or produced by the boat did not affect the measurements?
b) What was the ambient noise level at the presence of the bars? How was the acoustic power correlated with these morphological changes?
c) Was the signal sensitive to the dune's height change?
d) What was the bias of this signal towards the presence of gravel particles?

6. Final thought

Finishing up with the reading, I have a feeling that the comparison of the three methods is rather extensive and considering the length of modern scientific papers, it does not permit detailed analysis of each technique. However, this manuscript seems to have more informational character and not a comprehensive analysis of the deployed techniques.

Somehow the DTM method has a slightly different nature of processing and physical explanation. On the other hand, the current paper may involve fewer details about the techniques themselves and focus more on bedload dynamics.

In that sense and maybe in another paper, it could be very interesting to see the correlation between the corrected backscattering strength (see (Conevski, et al., (2020, accepted))) towards the acoustic power from the hydrophones and to focus on only on these two methods, as both represents "local" bedload measurements.

Specific comments (line by line)

L21-L23: Please correct the English. Something like: Although the parameters that control the self-generated noise still require analysis…

L23-24 The sentence does not contribute in the abstract and it is not clear. Does it mean that they are not able to measure if bedform migration is not happening?

L29: instead of: It constitutes a prerequisite to the accurate estimation please reformulate in: It is a prerequisite to an accurate estimation…

L: 41 : Please add the latest studies : (Conevski, et al., 2019) (Conevski, et al., (2020, accepted)) (Conevski, et al., 2020)

L 48: Are you talking about multi-beam?

L 64: more info about the PSD (or GSD). See general comment 1.

L 79: "below" substitute with Fig 2.

L 79 : "adopted here as reference". Maybe delete and start new sentence giving information that these data are used as reference data because this method so far is the most reliable… or something similar.

L 79: I suggest using a traditional or conventional method instead of classical.

L92-L100: See general comment 2. all sub comments.

L96: Would you upload a video of the high flood measurements? Thanks.

L102: You say two ADCPs. On Fig 2 I see only one and, in the description, I understood that only one was available per campaign.

L102 "positioned" could be replaced with installed.

L103: Which mode (IC-incoherent or SmartPulse HD?) was used in each case. Please add in the appendices (see, general comment 3 d.).

L120: Please use vector notation and see general comments(g.c.) (3 b.)

I will use g.c. as an abbreviation of general comment.

L124-L130: These are both kinematic models and essentially both describe a unit volume with its specific density moving with unit velocity. The first one is an assumption that the ADCPs are more sensitive to the particles and particle numbers and that by assuming characteristic particle Va would represent average in all 4 acoustic footprints. Considering g.c. 3c it is rather strange that you get better matching for Eq4 in Fig3b. But then the assumptions for Eq. 5 and the filtering are not correctly applied.

L143-L150: Did you get better results in these surveys, when ADCP was close to the bed?

L150: Large window like that would do simple de-spike low pass filter. Not the noise but the outliers.
L151: All negative values refer to the average values, if they were negative you discard from the dataset? Or you applied some kind of filtering on the raw data as given in g.c. 3a? If so please cite Conevski et al 2019.
L170: So why did you adopt 0.5 ?

L171: The accuracy of the echo sounding shell be very high. The problem is in the post-processing and identification in the bedforms, isn't it? See g.c 4a.

L174: What are the dimensions of the hydrophone?

L186: I assume the sampling was not synchronized?

L196: Is it the mean of all STDs?

L198: Again. Which of the two BTMA samplers were used in the average value?

L203-204: I am not sure if this comparison makes sense. To me, it looks that your RMA reg model could fairly represent the entire dataset (Rennie + yours, in the range of R2~0.5). Would be interesting to see. Check g.c. 3e.

L210-211: OK now I see that the average negative values were extracted not as in Conevski et al. 2019.

L213: Why did not you calculate for each point separately? What does constant means? For all data? If this is the case then it makes sense eq 4 to perform better.

L219-220> Could you clarify this sentence. Do you mean Eq5 is better than Eq4 or the opposite? Eq.4 does not assume any active layer thickness, but average particle size.

Fig3b.: Do you have something to say about the points down in the middle of the Fig3b?

L221-230: But that is reasonable. See g.c. 5a-b

L243: Does it involve the bedload concentration or just the velocity?

L254: The comparison is not consistent could you please clarify. The apparent velocity measures the velocity top layer or dynamical active layer, whereas the dune celerity is the mobility of the exchange event active layer, according to Church (2019).

L271: Perhaps, sampling sounds better than gauging.

L290: It is impressive if the DTM method somehow worked out in this morphology.

Fig8. Was this data used in the calibration of the Eq14 and 15 as well?

L303: bedload axis. Do you refer to a bedload active width?

L313-L317 +Fig9: Did you use Eq14 to calculate the transport rate using the raw apparent velocities, estimating qs for every ensemble of va?
If this is the case, the Eq14 is not calibrated over instantaneous values but over long averages, so It is not statistically sustained. Why don't you use only va?
L325: remove the comma.

Fig10: same as the comment above. The equation is not calibrated for the instantaneous values of the instrument.

L325-328: Do you suggest that the hydrophones are sensitive to the bar formation??

L352: "bedload transport occurs over a very low thickness", please reformulate to: in weak bedload transport conditions the BTMA sampler most likely performed with reduced efficiency … and cite literature.
L365: "during few time" to "little time" or

L378-384: This is very superficial information given in a way to suppress the ADCPs under the hydrophones. And it is only partially true. Please correct according to the general comments 3 and cite accordingly.
L378: (Conevski, et al., 2019)+ (Conevski, et al., (2020, accepted))+ (Conevski, et al., 2020)

L380: (Conevski, et al., 2020)

L385: Instead of "water depth heterogeneity" please use beam footprint heterogeneity

L386: After the filtering (Conevski, et al., 2019) negative values should not exist. But if there is beam heterogeneity this adcp value is rather incorrect and shell be eliminated not used as null. Also if somehow all the beams are sampling the dunes trough (lee side) the measurement is also incorrect if one aims to estimate the bedload transport over a cross-section.

L392: Please order the citations.

L392-395: This is not true. See my other comments. Beside active layer with a height of D50 would assume only rolling of the particles.
L395: You have observed only rolling of particles? Please send one video. I could only imagine this in incipient motion, although It is also impossible. On the contrary, (Lajeunesse, et al., 2010) and (Nin˜o, et al., 1998) report that most of the particles are actually jumping, meaning at least 2D50.

L396: Please do not confuse suspension or suspended load with bedload and its active layer. Saltation is not a suspension.
L397-404: This should be reformulated because it is not true. Please see the general comments regarding the ADCP and the comment above. And cite (Conevski, et al., 2020) for the 3MHz

L408: Indeed, physical samplers sample the dynamical active layer, thus more comparable to the hydrophones and adcps. The usual approach is to integrate the sampled bedload t.rate over the active width and compare to the DTM.

L421: This is very subjective. Conevski et al 2020 report 5min sampling per position using the M9. So in total 30 min, although no bars are presented in the study.

L440: Could you elaborate on how this is not valid for the hydrophones if they are in a fixed position? In the same way, by use of the RTK one could move the ADCPs and estimate the BT velocity, or just subtract the boat velocity from the BT velocity measured by another device or the boat itself. Or if it is certain that one dune has passed below the ADCPs then using the filtering (Conevski, et al., 2019) the data shell be fairly accurate.

L443: water depth not flow depth.

L452-456: Could you please reformulate. Do not really understand how the hydrophones are sensing the bars. I assume you mean change of the dunes which is just changing of the instantaneous transport and the shape of the dune… It could be another structure, not necessarily a bar… Is this correct?

L459: Please change Conversely to on the contrary.

L462: This is also partially true. The accuracy is within this range, but this stands for moving boat measurements. The laboratory data suggest lower values until 1mm/s check (Conevski, et al., 2020).

L469-L471: Please correct according to the general comments for the ADCP. If **strongly** correlated wit the grain size, please show the correlation.

References

Banhold Karin [et al.] Underestimation of sand loads during bed-load measurements—a laboratory examination [Conference]. - 2016. - pp. 1036-1041.

Brumley B. H. [et al.] Performance of a broad-band acoustic Doppler current profiler [Journal] // IEEE Journal of Oceanic Engineering. - 10 1991. - Vol. 16. - pp. 402-407. - ISSN: 0364-9059.

Conevski S. [et al.] Acoustic sampling effects on bedload quantification using acoustic Doppler current profilers [Journal] // Journal of Hydraulic Research. - 2020. - pp. 1-19.

Conevski S. [et al.] Laboratory investigation of the apparent bedload velocity measured by ADCPs [Journal] // Journal of Hydraulic Engineering. - 2019.

Conevski S. [et al.] TOWARDS AN EVALUATION OF BEDLOAD TRANSPORT CHARACTERISTICS BY USING DOPPLER AND BACKSCATTER OUTPUTS FROM ADCPS [Journal] // Journal of Hydraulic Research. - (2020, accepted).

Conevski Slaven [et al.] Bedload Velocity and Backscattering Strength from Mobile Sediment Bed: A Laboratory Investigation Comparing Bistatic Versus Monostatic Acoustic Configuration [Journal] // WATER. - 2020. - 3318 : Vol. 12.

Curran Joanna C. and Wilcock Peter R. Characteristic dimensions of the step-pool bed configuration: An experimental study [Journal] // Water Resources Research. - 2005. - Vol. 41.

Frings Roy M and Stefan Vollmer Guidelines for sampling bedload transport with minimum uncertainty [Journal] // Sedimentology. - 2017.

Gaeuman D. and Jacobson R. B. Acoustic bed velocity and bed load dynamics in a large sand bed river [Journal] // J Geophys Res Earth Surf. - 2006. - F02005 : Vol. 111(F2).

Gaweesh Moustafa and C. van Rijn Leo Bed-Load Sampling in Sand-Bed Rivers [Journal] // Journal of Hydraulic Engineering-asce - J HYDRAUL ENG-ASCE. - 12 1994. - Vol. 120.

Guerrero M., Lamberti, A. Flow field and morphology mapping using ADCP and multibeam [Journal] // J. Hydraulic Eng.. - 2011. - 137(12). - pp. 1576-1587.

Lajeunesse E., Malverti L. and Charru F. Bed load transport in turbulent flow at the grain scale: Experiments and modeling [Journal] // Journal of Geophysical Research: Earth Surface. - 2010. - 4 : Vol. 115.

Nin˜o Y and Garcı`a M Experiments on saltation of sand in water. [Journal] // J Hydr Eng. - 1998. - Vol. 124. - pp. 1014–1025.

Parker Gary ID SEDIMENT TRANSPORT MORPHODYNAMICS with applications to RIVERS AND TURBIDITY CURRENTS [Book]. - [s.l.] : St. Anthony Falls Laboratory, Mississippi River at 3rd Avenue SE, 2004.

R. Lee Gordon RD Instruments Acoustic Doppler Current Profiler Principles of Operation A Practical Primer [Report]. - San Diego, California : RD Instruments, 1996.

Rennie C. D., Millar R. G. and Church M. A. Measurement of bed load velocity using an acoustic doppler current profiler [Journal] // Journal of Hydraulic Engineering. - 2002. - 128 : Vol. 5. - pp. 473-483.

Rennie C.D and Villard P. V. Site specificity of bed load measurement using an acoustic Doppler current profiler [Journal] // J. Geophys. Res.. - 2004. - F03003 : Vol. 109.

Rennie C.D. [et al.] "Calibration of aDcp apparent bedload velocity to bedload transport rate" in Gravel-Bed Rivers [Book Section] // Processes and Disasters. - [s.l.] : Tsutsumi, D., and Laronne, J., Wiley., 2017.

Shiel F. Douglas Aquatic Habitat Bottom Classification Using ADCP [Journal] // JOURNAL OF HYDRAULIC ENGINEERING. - 2010. - pp. 336-342.

Urick Robert J. Principles of Underwater Sound [Book]. - [s.l.] : McGraw-Hill, 1983.

Wilcock Peter R., Kenworthy Stephen T. and Crowe Joanna C. Experimental study of the transport of mixed sand and gravel [Journal] // Water Resources Research. - 2001. - Vol. 37. - pp. 3349-3358.